

Nem-paraméteres (eloszlásmentes) statisztikai módszerek 2017

Makara Gábor

A terminológia magyarázata

- A paraméteres próbák a **normális (vagy valamelyik más) eloszlás** paraméterei (tulajdonságait) használják: pl. várható érték és szórás
- A nem-paraméteres eljárások nem használják ezeket a konkrét eloszlás paramétereket
- Az **eloszlásmentes** terminológia: ne vezessen rossz értelmezésre!

Megjegyzések

- A nem-paraméteres próbák esetében nem kell teszt arra, hogy teljesül-e a minták eloszlásának normalitása vagy valamilyen ismert eloszláshoz illeszkedése
- U próbák: elnevezés, ha a próbastatisztika normális eloszlású
- A kutatás során vizsgált beavatkozás/eltérés hatása lehet az eloszlás eltolása , az eloszlás más paraméterének/tulajdonságának változása (szórás, ferdeség), az eloszlás típusának megváltozása is
 - Például egy sok tényező által befolyásolt állapot helyett egy-két tényező által meghatározott állapot keletkezik

Az előadás céljai

- Módszerek általános áttekintése
- Módszerek alkalmazásának optimalizálása
- A módszerek bemutatása az egyszerűbbtől az összetettebbig
- Előjel teszt és Wilcoxon előjeles rangteszt alkalmazása
- Mann-Whitney U teszt alkalmazása
- Kitekintés nem-paraméteres többes összehasonlításokra
 - Kruskal-Wallis, Friedman tesztek
- Módszerválasztás elvei és gyakorlata

Hat-e a tesztoszteron a hím patkány agressziójára

az adatokat rendeztem a harapások száma szerint, a könnyebb áttekinthetőségért (megtekintés, explorálás...)

Sorszám	Kontroll – harapások száma	Tesztoszteron – harapások száma
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	2
8	0	2
9	1	2
10	1	3
11	1	4
12	1	5
13	1	6
14	1	7
15	1	8
16	3	9
17	3	26
18	4	27
19	9	

Dilemmák

- Az agresszió mértéke folytonos változó-e?
- A harapások száma milyen változó? Folytonos vagy diszkrét?
- Sok az azonos harapásszám, mit tegyünk ezekkel?
- Milyen eloszlást követhetnek az adatok?

Elemzési lehetőségek – melyik lesz jó?

- t próba, 2 mintás
- Welch próba, d próba
- transzformáció
- Osztályokba sorolás: pl. szelíd – harapós – vad
- Két osztályba sorolás:
 - a közös medián alatt/felett hány megfigyelés, 2x2-es táblává redukáljuk
- Párba állítás, melyik harap többet? (véletlen párosítással)
- Sorba állítás – rangsorolás
 - Véletlenszerűen keverednek?
 - Egyik csoport van a közös rangsor elején (vagy végén)?

Feladatok lehetnek:

- Eloszlásfüggvény (sűrűségfüggvény) becslése (illeszkedés vizsgálat) (empirikus, minta eloszlás hasonlítása elméleti eloszláshoz)
- Eloszlás jellemzők becslése (várható érték, medián, szórás, szimmetria/ferdeség, stb.)
- Hipotézis vizsgálat

A kutatás lépései vázlatosan

1. Előkészületek (szakmai, előkísérlet, mérési skála, explorálás: várható szóródás, várható hatásméret, stb.)
2. Hipotézisek megfogalmazása (egy vagy kétoldalú), az első fajú (kritikus érték - alfa) és a második fajú hiba meghatározása – részei a tervezésnek
3. A detektálni kívánt hatás mérete (SD értékekben), a másodfajú hiba és a szükséges elemszám meghatározása (paraméteres módszerrel)
4. A megfelelő teszt és tesztstatisztika kiválasztása
5. Döntési szabály: mikor vetjük el a H_0 -t, mikor nem vetjük el
6. Elvégezzük a kutatási adatgyűjtést
7. Kiszámítjuk a tesztstatisztikát
8. A tesztstatisztika értékét mérlegeljük, hasonlítjuk a döntési szabályhoz
9. Megfogalmazzuk a teszt alapján jogos statisztikai és szakmai következtetéseket
10. Nem folytatjuk az adatgyűjtést, ha a kimenetel nem szignifikáns. Új vizsgálatra kerülhet sor

A hipotézis vizsgálat kimenetele

Döntési táblázat

	H_0 igaz	H_0 hamis
H_0 elfogadása	Helyes döntés ($1-\alpha$)	Másodfajú hiba (β hiba)
H_0 elutasítása	Első fajú hiba (α hiba)	Helyes döntés (Power = $1-\beta$)

A döntési küszöbök értékei

- Elsőfajú hiba (α , alfa), második fajú hiba (β , béta)
- A másodfajú hiba meghatározásához kell a **hatásméret** (d) és az **elemszám** (N)
- A nem paraméteres módszereknél a β meghatározása nehéz
- a “power”, $1-\beta$, a módszer ereje is gyakran nehezen, gyakran modellezéssel meghatározható
- **Az optimális próbát kell megkeresnünk!**
 - A próbák egymáshoz hasonlíthatóak, szimulációval, nagy az irodalma
- Egy vagy kétoldali próbát is végezhetünk, ez a döntés szakmai

Kiinduló feltételezések

- A mért változó
 - nominális skálán
 - ordinális skálán (rangskálán)
 - numerikus skálán (diszkrét, vagy folytonos)
(eloszlása nem standard-normális)
- A null hipotézis
 - eloszlások azonossága
 - a mediánok azonossága (medián eltolása?)
- A minták száma
 - Lehet 1, 2, >2

Ha véletlen hiba nem normális eloszlású

- Több úton kereshetjük a megoldást
 - transzformálással „normalizáljuk” az eloszlást
 - megvizsgálhatjuk, vajon a módszer kellően robusztus-e?
 - Használható lehet, ha a normális eloszlástól eltérés „nem nagy”
 - Ha igen alkalmazzuk a szuboptimális módszert..
 - „Eloszlás-mentes” másszóval „nem-paraméteres” módszereket alkalmazunk (rendezett minták elméletét alkalmazzuk)
 - **folytonos eloszlásfüggvényt** feltételezünk, azaz azt, hogy az összes mintaelem 1,0 valószínűséggel különbözik

A tesztek fontos tulajdonságai

Név	Mérési skála	Eloszlás	Megkötések, tulajdonságok	Csoportok
Kontingencia tábla	nominális	Irreleváns		2 vagy sok független
Előjel próba	Legalább rangskála	Folytonos	Döntetlen kihagyása	2 - párosított
Wilcoxon párosított teszt	Legalább rangskála	Folytonos	Döntetlen kihagyása	2 – párosított
Friedman próba	Legalább rangskála		Ismételt mérés, randomizált blokk	2 vagy több összefüggő
Mann-Whitney U teszt	Legalább rangskála	Folytonos	Mediánok különbsége	2 – független
Wald-Wolfowitz run teszt	Legalább rangskála	Folytonos	Érzékeny az eloszlás alakjára	2 – független
Kolmogorov-Szmirnov D teszt		Tetszőleges	Érzékeny az eloszlás alakjára	2 – független
Medián teszt	Legalább rangskála		Visszavezet a 2xk-s táblára	2 – vagy sok független
Kruskal-Wallis próba	Legalább rangskála	Folytonos	Kettőnél több csoport	Több mint 2

Rang transzformáció

- A megfigyeléseket nagyság szerint sorba állítjuk
- A megfigyelések helyett vesszük a sorszámukat=rangszám
- A rangokból számolhatunk rangstatisztikákat
- A rangstatisztikák invariánsak a minta elemek minden szigorúan monoton transzformálására.
- A rangszámok **diszkrét és egyenletes eloszlású** változók
- A $[0,1]$ intervallumban egyenletes eloszlású, rendezett minta rangszámaira, a **k-adik mintaelemre**:
 - átlag: $k/(n+1)$
 - szórás: $k(k+1)/(n+1)(n+2)$
 - szórásnégyzet: $k(n-k+1)/(n+1)^2(n+2)$

Összefüggő minták nem-paraméteres tesztjei

- Terminológia: párosított, összefüggő, kapcsolt, ismételt
- A párosítás a kutatási terv része
 - Azonos alanyon több mérés
 - Különböző alanyokon párban (csoportosítva) történő kezelés és mérés
 - Általában előnyösebb, mint az azonos elemszámú nem-párosított vizsgálat
- Előjel teszt
 - A párban mért eredmények különbségének előjele
- Wilcoxon féle előjeles rangteszt
 - Párban mért eredmények különbsége, rangtranszformálva, előjel figyelembe véve
- A párosítás kiterjeszthető a többes összefüggésre is
 - Azonos alanyon több mérés
 - Különböző alanyokon azonos körülmények között több kezelés és mérés

Példa párosított mintákra, rendezve:
Noradrenalin hatása a máj nyirok NEFA tartalmára

NEFA koncentráció $\mu\text{eq}/\text{l}$ NE <u>előtt</u>	NEFA koncentráció $\mu\text{eq}/\text{l}$ NE <u>után</u>	Különbség	Előjel	Különbség abszolút értékének rang száma	Előjeles rang szám
784	680	-104	-	1	-1
1114	1448	334	+	2	2
945	1340	395	+	3	3
780	1182	402	+	4	4
1077	1488	411	+	5	5
780	1195	415	+	6	6

Páros tesztek – előjel teszt

- Egy minta esete, amikor egy igen/nem, vagy +/- vagy A/B sorozatból áll a mérési minta. H_0 : az egyik és másik előjel $p=0,5$ valószínűséggel fordul elő.
- Párosított minták, amelyekben van előtte/utána, vagy a pár 1. és 2. elemére olyan mérés, amelynek eloszlása nem ismert, de folytonos és ordinális skálán értelmezett.
- A tesztet visszavezethetjük az egy minta esetére, úgy hogy a párban mért értékek különbségének előjelét vesszük mérési adatnak.
- Ha egy párban azonos értéket mérünk mindkét esetben, akkor a különbség nulla és az ilyen eseteket kihagyjuk az elemzésből, csökkentve az elemszámot is.
 - A változó folytonosságából következően elméletileg nem lehet két azonos érték, de a gyakorlatban előfordul, hogy nem lehet különbséget tenni.

Előjel próba táblázata

K	n	n	n	n	n	n	n	n	n	n
	1	2	3	4	5	6	7	8	9	10
0	0.5	0.25	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001
1	0.5	0.50	0.375	0.250	0.156	0.094	0.055	0.031	0.018	0.010
2		0.25	0.375	0.375	0.313	0.234	0.164	0.109	0.070	0.044
3			0.125	0.250	0.313	0.313	0.273	0.219	0.164	0.117
4				0.063	0.156	0.234	0.273	0.273	0.246	0.205
5					0.031	0.094	0.164	0.219	0.246	0.246
6						0.016	0.055	0.109	0.164	0.205
7							0.008	0.031	0.070	0.117
8								0.004	0.018	0.044
9									0.002	0.010
10										0.001

Ahol:

n a megfigyelések száma

k a pozitív (vagy a negatív) előjelek száma

p a táblázatban feltüntetett számok

A piros számok a szignifikáns p (valószínűség) értékeket jelzik (kétoldalas próba !)

Az előjel próba ereje (érzékenysége)

- A táblázat adatai mutatják, hogy az $N=5$ esetben az előjel próba kimenetének valószínűsége sem érheti el a $p=0,05$ -ös küszöbértéket.
- $N=6$ -tól $N=8$ -ig csak egynemű előjelek adnak szignifikáns eredményt, $N=9$ -től kezdve már az egynemű mellett egy kisebbségben lévő előjel mellett is lehet szignifikáns a próba eredménye.
- A **próba ereje** számítható, meghatározott alternatív hipotézis esetére, ahol H_0 a $p=0,5$ valószínűsége az egyik előjel előfordulásának, az alternatív hipotézis pedig a $p \neq 0,5$ konkrét esete.
- Intuitíve a próba ereje annál nagyobb, minél távolabb esik az alternatív hipotézisben (H_A) megadott valószínűség a $0,5$ -től
 - A hatás mérete a $|p_A - 0,5|$, az egyik előjel előfordulási valószínűségének detektálni kívánt eltérése a H_0 -tól.

Normal közelítés az egyoldalú előjel teszt erejére, különböző mediánokra, $N = 16$ esetére

theta	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,90
Erő (Power)	0,0461	0,0918	0,1629	0,2639	0,3960	0,5546	0,7255	0,8802	0,9773

Table 4.1 Gibbons, J Nonparametric statistical inference , Dekker, 2003

Páros tesztek: Wilcoxon előjeles rang teszt

- A mért változó folytonos és legalább ordinális skálán mért
- A mért változó eloszlása szimmetrikus a medián körül
- A páros mérések esetében a különbségek is ordinális skálán mérhetőek
- A null hipotézis a minták mediánjainak azonossága
 - Ebben az esetben H_0 : pozitív rangok összege (T^+) = |negatív rangok összege (T^-)|
 - A tesztstatisztika a rangokból számolt rangösszeg (T), a kettő közül a kisebb
- A próba aszimptotikus relatív hatékonysága (efficiency) - ARE
 - $\geq 0,864$ minden szimmetrikus eloszlásra, amelynek mediánja = 0
 - ≈ 0.955 , közelíti az páros t teszt erejét, ha a minták eloszlása normális
 - ARE: azon elemszámok aránya, amely azonos próbaerőt ad.

A Wilcoxon féle előjeles rang teszt ereje

- A teszt erejét szimulációval lehet (könnyebben) megállapítani
- Például $N=13$, $\alpha=0,05$, $M_0=0,5$ és $M_1=0,6256$ (alternatív hipotézis)
 $T^+ \geq 70$ -nél $\alpha=0,045$
- A normál eloszlásból generálnak 1000 random mintát ezekkel a paraméterekkel, és mindegyikre meghatározzák a Wilcoxon próba eredményét
- Az elvetett H_0 esetek számát elosztva 1000-el kapunk egy becslést a próba erejére.

Példa párosított mintákra, rendezve:
Noradrenalin hatása a máj nyirok NEFA tartalmára

NEFA koncentráció $\mu\text{eq}/\text{l}$ NE <u>előtt</u>	NEFA koncentráció $\mu\text{eq}/\text{l}$ NE <u>után</u>	Különbség	Előjel	Különbség abszolút értékének rang száma	Előjeles rang szám
784	680	-104	-	1	-1
1114	1448	334	+	2	2
945	1340	395	+	3	3
780	1182	402	+	4	4
1077	1488	411	+	5	5
780	1195	415	+	6	6

A máj nyirok NEFA változása NA hatására

Data: T-test for Dependent Samples (NEFA) (10v by 2c)

T-test for Dependent Samples (NEFA)
Marked differences are significant at $p < ,05000$

Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p	Confidence -95,000%	Confidence +95,000%
Előtt	913,333	155,1382								
Után	1222,167	293,8873	6	-308,833	204,3912	-3,70116	5	0,013983	-523,329	-94,3379

Data: Wilcoxon Matched Pairs Test (NEFA) (4v by 1c)

Wilcoxon Matched Pairs Test (NEFA)
Marked tests are significant at $p < ,05000$

Pair of Variables	Valid N	T	Z	p-value
Előtt & Után	6	1,000000	1,991741	0,046400

Data: Sign Test (NEFA) (4v by 1c)

Sign Test (NEFA)
Marked tests are significant at $p < ,05000$

Pair of Variables	No. of Non-ties	Percent $v < V$	Z	p-value
Előtt & Után	6	83,33333	1,224745	0,220671

Előjel próba táblázata

K	n	n	n	n	n	n	n	n	n	n	n
	1	2	3	4	5	6	7	8	9	10	
0	0.5	0.25	0.125	0.063	0.031	0.016	0.008	0.004	0.002	0.001	
1	0.5	0.50	0.375	0.250	0.156	0.094	0.055	0.031	0.018	0.010	
2		0.25	0.375	0.375	0.313	0.234	0.164	0.109	0.070	0.044	
3			0.125	0.250	0.313	0.313	0.273	0.219	0.164	0.117	
4				0.063	0.156	0.234	0.273	0.273	0.246	0.205	
5					0.031	0.094	0.164	0.219	0.246	0.246	
6						0.016	0.055	0.109	0.164	0.205	
7							0.008	0.031	0.070	0.117	
8								0.004	0.018	0.044	
9									0.002	0.010	
10										0.001	

Ahol:

- n a megfigyelések száma
- k a pozitív (vagy a negatív) előjelek száma
- p a táblázatban feltüntetett számok

A piros számok a szignifikáns p (valószínűség) értékeket jelzik (kétoldalas próba !)

Friedman próba

- A blokk hatás kiküszöbölésére, több mint 2 csoport esetén.
- Modell: $X_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$
- itt a β_j a kiküszöbölendő blokk „hatás”
- A $\sum \tau_i = 0$ és a $\sum \beta_j = 0$ összefüggések teljesülnek
- $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \dots = \tau_n$
- $H_1: \tau_i$ nem mind egyenlők
- Eljárás: minden blokkban külön rendezzük az adatokat, és helyettesítjük őket a rangszámokkal.
- S statisztikát számolunk, ami X^2 eloszlású

Két független csoport összehasonlítása

Feltételezések

- Folytonos eloszlás és
- legalább ordinális skála és
- független megfigyelések

Általánostól a specifikus felé haladva

1. Wald-Wolfowitz próba
2. Kolmogorov-Szmirnov teszt
3. Mann-Whitney próba

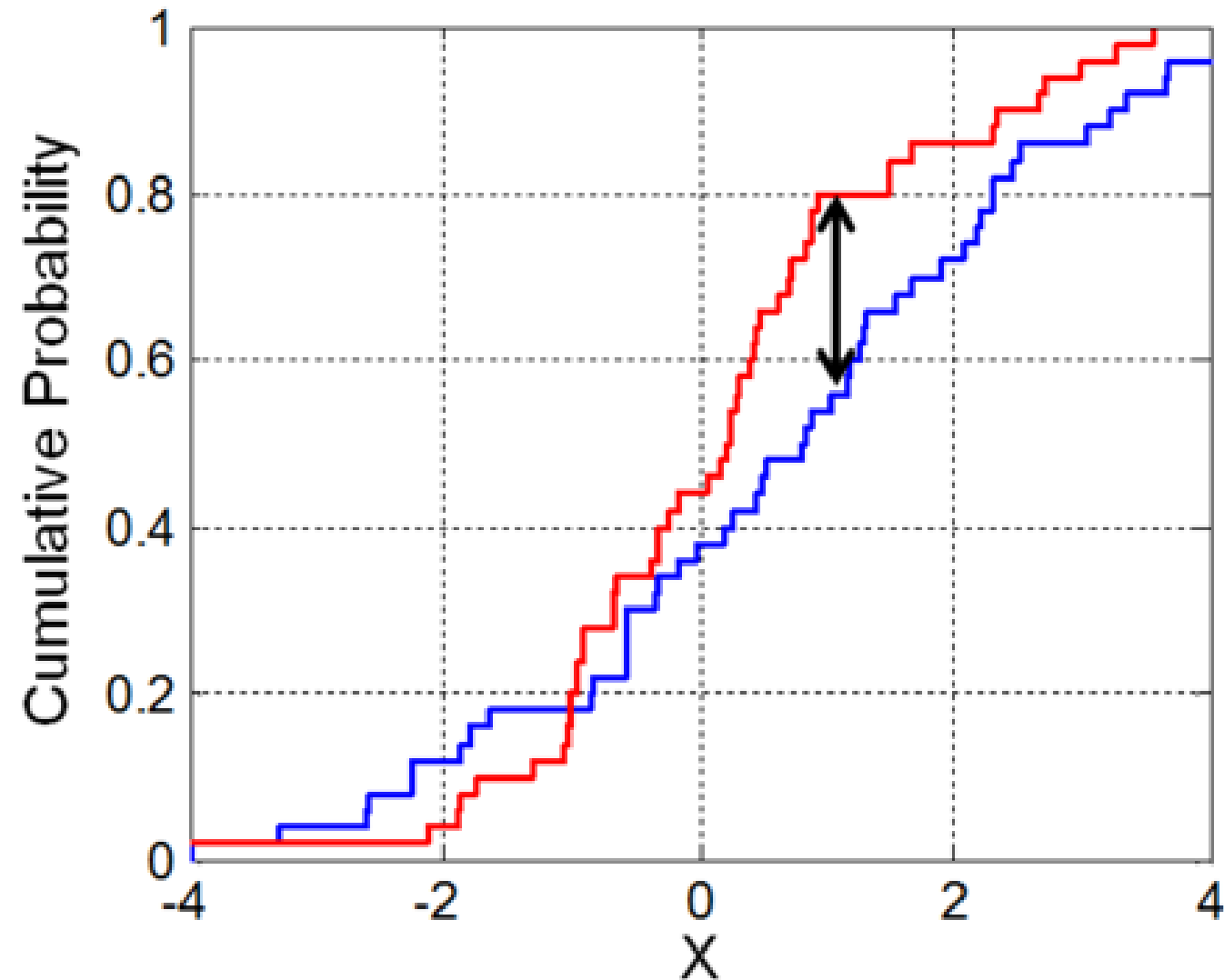
Wald-Wolfowitz sorozat próba (angolul „runs” test)

- Két véletlen minta, függetlenek, folytonos eloszlásból
- H_0 : a két minta azonos eloszlásból származik
 H_A : a két minta különböző eloszlásokból
- A két minta egyesítve rendezése után a **keveredést vizsgáljuk**: az egyes mintaelemek véletlenszerűen követik egymást, azonos mintából álló (homogén) sorozatok (blokkok) hosszából következtetünk arra, hogy a két minta azonos eloszlásból származik-e?
- A sorozatok száma a teszt statisztika, sem a nagy számú sorozat, sem a kis számú sorozat nem véletlenszerű. (példák: babababab vagy aaaabbbbb, 8-ból 4 vagy 2)
- A teszt nagyon általános!
A próba ereje nehezen kezelhető, kellene hozzá specifikus alternatív hipotézis.
- Kevésbé általános próbák hasznosabbak és hatékonyabbak lehetnek
- Elsősorban előzetes vizsgálatokban hasznos, amikor még kevés ismeretünk van a vizsgálat tárgyáról

Kolmogorov-Szmirnov teszt

- Két minta empírikus eloszlását (görbét) hasonlítja össze
- Folytonos eloszlás, független random minták
- H_0 : a két minta azonos eloszlásból származik
 H_A : a két minta különböző eloszlásból származik
- A teszt-statisztika a maximum abszolút különbség: $D_{m,n}$
- A teszt az eloszlások sokféle tulajdonságára érzékeny
- Használata elsősorban az előzetes vizsgálatoknál indokolt.
- A próba ereje általában nagyobb, mint a Wolf-Wolfowitz sorozatpróbájé

Két empirikus eloszlásfüggvény távolsága



Mann-Whitney U teszt – Wilcoxon rangösszeg teszt

- Alapelve: az egyesített adatok növekvő sorrendjében a minták értékeinek nagyságából
- Folytonos eloszlás, két random, független minta
- U statisztika: Az eseteknek száma, amelyben az adott csoport eleme megelőzi a másik csoport elemét. Egy precedencia statisztika, hasonlít a Wald-Wolfowitzra
- H_0 : a két eloszlás mediánjai azonosak
 H_A : a két eloszlás mediánjai eltérnek
- Mediánok különbségére különösen érzékeny
- A próba ereje és a hozzá szükséges minta elemszám számolható adott medián-eltoláshoz, de nincs jelen a Statistica v13-ban.
- A Student t teszthez mért relatív hatékonysága $ARE \geq 0,864$ és a normális eloszlás esetében az aszimptotikus relatív hatékonyság: 0,9550
- Lényegesen jobb is lehet a Student t tesztnél, erősen ferde eloszlásoknál.

Mann-Whitney teszt (folytatás)

- A legjobb nemparaméteres módszer a minták eltolásának esetére
- A próba ereje egy megadott eltolásra (mediánok különbségére) számolható
- A **Wilcoxon féle rangösszeg teszt** mediánok különbségére
 - Ekvivalens a Mann Whitney U teszttel
 - Ekvivalens a rangokon végzett ANOVA-val is.
- Ha az egyesített adatok közös rangtranszformációja után összegezzük az egyes minták rangszámait, akkor - az összeg nagyobb lesz, annál a mintánál, amelyik mediánja nagyobb.

Rangösszeg próba

	Tumor méret változása (cm)							
Kezelt	1,29	1,60	2,27	1,31	1,81	2,21		
Nem kezelt	0,96	1,14	1,59					

	Tumor méret változása (cm)									W
Kezelt			1,29	1,31		1,60	1,81	2,21	2,27	
Nem kezelt	0,96	1,14			1,59					8
Rangszenek	1	2	3	4	5	6	7	8	9	

Azonos rangszámok kezelése

- Az elmélet szerint a folytonos változók esetében nem lehetséges két azonos mért érték
- A gyakorlatban a folytonos és ordinális skálán mért változóknál is előfordul több azonos mért érték, melyek a korlátozott mérési pontosság miatt lehetnek azonosak. Ezeknél eldöntetlen, hogy melyik kapjon kisebb és melyik nagyobb rangszámot.
- A döntetlen esetek kezelésének leggyakoribb esete, hogy az ezekre jutó rangszámokat a rangszámok átlagával helyettesítjük.
- A következő adattáblában a nem-egész (tizedesvesszőt tartalmazó) számok ilyen döntetlen esetekhez rendelt rangszám értékek
- A döntetlen esetek kezelhetők úgy is, hogy az azonos mért értékek esetében sorsoljuk a rangszám kiosztást, véletlenszerűen ítéljük oda az egyes csoportokhoz a mérések alapján eldöntetlen rangokat

Sorszám	Kontroll - harapás	Rangszám	Tesztoszteron - harapás	Rangszám
1	0	7,50	0	7,50
2	0	7,50	0	7,50
3	0	7,50	0	7,50
4	0	7,50	0	7,50
5	0	7,50	0	7,50
6	0	7,50	0	7,50
7	0	7,50	2	23,00
8	0	7,50	2	23,00
9	1	18,00	2	23,00
10	1	18,00	3	26,00
11	1	18,00	4	28,50
12	1	18,00	5	30
13	1	18,00	6	31
14	1	18,00	7	32
15	1	18,00	8	33
16	3	26,00	9	34,50
17	3	26,00	26	36
18	4	28,50	27	37
19	9	34,50		

A tesztoszteron hatás elemzése Statistica v13-ban

Data: T-tests; Grouping: KEZELES (vereked) (11v by 1c)											
T-tests; Grouping: KEZELES (vereked) Group 1: Kontroll Group 2: Teszto-s											
Variable	Mean Kontroll	Mean Teszto-s	t-value	df	p	Valid N Kontroll	Valid N Teszto-s	Std.Dev. Kontroll	Std.Dev. Teszto-s	F-ratio Variances	p Variances
AGGR_N	1,368421	5,611111	-2,18774	35	0,035456	19	18	2,191157	8,153952	13,84807	0,000001

Data: Median Test, Overall Median = 1,00000; AGGR_N (vereked)* (3v by 7c)						
Median Test, Overall Median = 1,00000; AGGR_N (vereked) Independent (grouping) variable: KEZELES Chi-Square = 7,835840 df = 1 p = ,0051						
Dependent: AGGR_N	Kontroll	Teszto-s	Total			
<= Median: observed	15,00000	6,00000	21,00000			
expected	10,78378	10,21622				
obs.-exp.	4,21622	-4,21622				
> Median: observed	4,00000	12,00000	16,00000			
expected	8,21622	7,78378				
obs.-exp.	-4,21622	4,21622				
Total: observed	19,00000	18,00000	37,00000			

A tesztoszteron hatás elemzése Statistica v13-ban

Data: Wald-Wolfowitz Runs Test (vereked)* (10v by 1c)										
Wald-Wolfowitz Runs Test (vereked) By variable KEZELES Marked tests are significant at $p < .05000$										
Variable	Valid N Kontroll	Valid N Teszto-s	Mean Kontroll	Mean Teszto-s	Z	p-value	Z adjstd	p-value	No. of Runs	No. of ties
AGGR_N	19	18	1,368421	5,611111	-2,16462	0,030417	1,997767	0,045743	13	8

Data: Kolmogorov-Smirnov Test (vereked)* (9v by 1c)									
Kolmogorov-Smirnov Test (vereked) By variable KEZELES Marked tests are significant at $p < .05000$									
variable	Max Neg Differnc	Max Pos Differnc	p-value	Mean Kontroll	Mean Teszto-s	Std.Dev. Kontroll	Std.Dev. Teszto-s	Valid N Kontroll	Valid N Teszto-s
AGGR_N	-0,456140	0,00	$p < .05$	1,368421	5,611111	2,191157	8,153952	19	18

Data: Mann-Whitney U Test (w* (10v by 1c)										
Mann-Whitney U Test (w/ continuity correction) (vereked) By variable KEZELES Marked tests are significant at $p < .05000$										
variable	Rank Sum Kontroll	Rank Sum Teszto-s	U	Z	p-value	Z adjusted	p-value	Valid N Kontroll	Valid N Teszto-s	2*1sided exact p
AGGR_N	301,0000	402,0000	111,0000	-1,80802	0,070605	-1,86658	0,061962	19	18	0,070274

A tesztoszteron hatás statisztikai tesztelésének értelmezése

- A két mintás t próba használata ebben az esetben szakszerűtlen.
Főbb okai:
 - A harapások száma erősen ferde eloszlású, távol áll a normálistól
 - A szórások különbözőségére az F próba p értéke nagyon kicsi, nem teljesül a szórások azonosságának feltétele
- A Wolf-Wolfowitz és a Kolmogorov-Szmirnov tesztek szignifikánsak, jelezve, hogy a két minta nem egy eloszlásból származik.
- A Mann-Whitney U próba esetében a null hipotézisre a p érték 0,06 vagy 0,07 körül van, aminek alapján a mediánok különbözősége nem tekinthető igazoltnak, további vizsgálatok lehetnének szükségesek
- Figyelembe vehetjük, hogy a szakirodalom szerint a tesztoszteron sok helyzetben fokozza az agresszivitást, az alternatív hipotézis már a vizsgálat előtt (a priori) valószínűbbnek volt tekinthető, mint a null hipotézis. Az adott vizsgálat után a szakmai következtetésben az előzetes valószínűséget és a vizsgálat által szolgáltatott bizonyítékokat együtt érdemes értelmezni

Kruskal-Wallis próba

- Kettőnél több minta nem-paraméteres összehasonlítása
- Modell: $X_{ij} = \mu + \tau_i + \xi_{ij}$
- $H_0 : \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = \dots = \tau_n$
- $H_1 : \tau_i$ nem mind egyenlők
- Az eljárás: minden megfigyelést együtt rangsorolunk, majd az eredeti adatok mellé írjuk csoportosítva a rangszámokat.
- A rangszám összegekből számoljuk a *H statisztikát*, aminek eloszlása $X^2(k-1, \alpha)$
- Ha a nullhipotézist elvetjük, többszörös összehasonlítást végezhetünk

A nem-paraméteres tesztek viszonya a paraméteres tesztekhez

Előny

- Szélesebb alkalmazási terület
- **Jobb próbaerő**, mint a paraméteres teszté, **ha a paraméteres feltételek nem teljesülnek.**

Hátrány

- **Gyengébb próbaerő**, mint a jól elvégezhető paraméteres teszt (feltételei teljesülnek)
- Számos elrendezéshez nincs elérhető optimális teszt
- A próba ereje és a szükséges **elemszám nehezen tervezhető** (gyakorlott statisztikus...)

Főbb üzenetek

- Optimális biostatistikai módszer:
függ az alkalmazott mérési skálától és az adatok eloszlásától
- Nem folytonos eloszlás
esetén korlátozottabbak a tesztválasztás lehetőségei
- A párosított minták esetében a módszerek érzékenyebbek
- A hatásméretre vonatkozó megfontolásokat itt is érdemes mérlegelni.
- A módszer erejét a megfelelő paraméteres tesztre könnyen számolhatjuk
- A reprodukálhatóság paraméterei egyelőre nem ismertek

A reprodukálhatóság krízise a tudományos szakirodalomban

- **A jelenség**
a tudományos szakcikkeknek töredékét sikerült utánvizsgálatokban reprodukálni.
- **Az okok**
 - Tervezési hibák, **alacsony elemszám és az ismeretlen „statisztikai erő”**:
 - **Értelmezési hibák: a null hipotézis, p érték és a konfidencia intervallum fogalmában**
 - Eredmények szelekciója, a negatív kutatási eredmények nem nyilvánosak
 - Rejtett (vagy eltírt) többszörös összehasonlítások
 - Adatok manipulálása „szignifikancia” elérésére:
- **A javítás módszerei**
 - A biostatisztikai oktatása színvonalának javítása:
 - **Hatásméret, elemszám, reprodukálhatóság szintjének tervezése:**
Mindig rögzíteni a vizsgálati protokollban:
 - adatrepresentativitás biztosítása
 - **egy vagy kétoldali próba használata**
 - **α -érték rögzítése**
 - **mintaszám meghatározás a vizsgálat előtt**
 - **alkalmazott statisztikai módszerek**
 - bootstrap-analízis, szimuláció
 - irodalom kutatás
 - pilot-study végzése.
 - **Adatmanipulálási szokások tárgyalása**
 - Megbízható negatív eredmények számára közlési lehetőség
 - Biostatisztikai módszertani leírás, biostatisztikai lektorálás
 - Tervezés lektorálása, nyilvánosan regisztrált kutatások elterjesztése
 - A reprodukálhatóság vizsgálatának tudományos értéként elismerése
 - Open Data közlés

(Bio)statisztikai hibák és orvoslásuk a jobb reprodukálhatóságért

- **Tervezés**

- **Hatásméret, megfelelő elemszám és a „statisztikai erő” (power)**
- A tervezési folyamatban szerepeljen a reprodukálhatóság becslése
- A szórás és **a hatás méret szerepeljen** a paraméterek között
- **A feltáró és a megerősítő vizsgálatok megkülönböztetése**
- Adattisztítás szabályainak rögzítése

- **Biostatisztikai szemlélet**

- **A hipotézis vizsgálat, a p érték és a konfidencia intervallum gondos értelmezése**
- A fals pozitívítási esély becslése, értelmezése
- **A szignifikancia szó használatának mellőzése**
- **Korrekt statisztikai módszerek használata**

- **Többszörös összehasonlítás a kutatási folyamatban:**

- Adathalászat sok mérés között
- A „szignifikáns” esetek feltüntetése tervezett vizsgálatként

- **Adat manipulálás módszerei „szignifikancia” elérésére**

- Kedvezőtlen adatok kihagyása a szórás csökkentésére
- **Menet közbeni esetszám növelés a $p < 0,05$ eléréséig**

- **A negatív kimenetelű kutatási eredmények kezelése**

- A pozitív eredmények közlésében a negatív vizsgálatok is megjelenítendőek
- A kutatási hipotézisek negatív eredményű tesztelésének közlése